

ОЦЕНИВАНИЕ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ ТЕКСТА ПРИ ПОМОЩИ НЕЧЁТКОЙ ЛОГИКИ

Поздняков М.В. (Университет ИТМО), доцент, кандидат технических наук Осипов Н.А. (Университет ИТМО)

Введение. Анализ больших массивов текстовых данных является важным направлением машинного обучения. Отдельную роль в обработке текста играют нейронные сети, которые существенным образом повышают качество решения стандартных задач классификации текстов и последовательностей, снижают трудоёмкость при работе непосредственно с текстами, а также позволяют решать новые задачи (например, создавать чат-боты). В то же время нейронные сети нельзя считать полностью самостоятельным механизмом решения лингвистических проблем, и они являются не единственным многозадачным математическим аппаратом [1, 2]. Для исследования рассуждений в условиях неопределенности, размытости, сходных с рассуждениями в обычном смысле, и поиска возможностей применения в вычислительных системах может применяться аппарат, основанный на нечёткой логике [3]. В рамках данного доклада будет представлен прототип системы, определяющей эмоциональную составляющую текста на основании модели, построенной на аппарате нечёткой логики.

Основная часть. Существует несколько этапов определения эмоциональной составляющей текста. Первым этапом является предобработка. Текст может содержать некоторое количество шума (лишние и пропущенные символы или буквы, ошибки правописания). В ходе предобработки происходит удаление ненужной информации, например URL, множественного повторения знаков, специальных символов. После этого данные проверяются на избыточность.[4] Для данной системы этап предобработки заключается в приведении всех слов к начальной форме и определении степени отрицательной или положительной окраски каждого слова. При определении эмоциональной составляющей слов учитываются отрицательные частицы. Коэффициенты каждого из слов суммируются, также подсчитывается количество восклицательных знаков и многоточий, после чего эти данные передаются на вход прототипу модели.

Разрабатываемая модель предназначена для определения эмоциональности отзыва, а также степени удовлетворённости товаром. Входные параметры: процент слов с положительной эмоциональной окраской, процент слов с отрицательной эмоциональной окраской, процент восклицательных знаков, процент многоточий, а также длина текста относительно самого длинного из выборки. Выходные параметры: степень эмоциональности отзыва и степень удовлетворённости пользователя товаром. В качестве основы была выбрана модель нечёткой логики. Такая математическая модель позволяет формализовать человеческие рассуждения, что уместно при оценке эмоциональной составляющей текста.[5]

Система была проверена на реальных данных. Для проверки модели были взяты 15 отзывов с сайта Яндекс.Маркет к смартфону Redmi Note 10 Pro: по 5 отзывов, соответствующих оценке 5/5, оценке 1/5 и оценке 3/5. В среднем значение удовлетворённости автора отзыва товаром было определено следующим образом: для отзывов на 5/5: 0,75, для отзывов на 3/5: 0,61, для отзывов на 1/5: 0,37

Выводы. Были рассмотрены существующие модели по извлечению эмоций. Был разработан прототип модели оценки эмоциональной окраски отзывов на основе математического аппарата нечёткой логики. Проверка модели на реальных данных подтвердила возможность применения аппарата нечёткой логики в задачах кластеризации отзывов покупателей в условиях размытости входных данных. Выяснилось, что параметры функций принадлежности для входных значений экспертами неосознанно завышаются.

Была проведена оптимизация данных значений в сторону смещения параметров в меньшую сторону.

Список использованных источников:

1. Глубинное обучение для автоматической обработки текстов [Электронный ресурс] — Режим доступа: <https://www.osp.ru/os/2017/02/13052221>, свободный. Дата обращения: 15.12.2021.
2. Есть ли альтернатива искусственным нейронным сетям? [Электронный ресурс] — Режим доступа: <https://postnauka.ru/faq/86374>, свободный. Дата обращения: 01.05.2022
3. Лепихова В.А. Ляшенко Н.В., Яковенко Е.А., Шестак С.Г., Рябоус А.Ю. Mathematical and software support for experimental work and algorithms for vibroacoustic diagnostics // Proceedings of SPIE - The International Society for Optical Engineering - SPIE, 2022. - Vol. 12251: Computer Applications for Management and Sustainable Development of Production and Industry (CMSD2021), 21-23 Dec. 2021, Dushanbe, Tajikistan. - Номер статьи 122510C. - 5 с. - 0.53 п.л.
4. Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. Current State of Text Sentiment Analysis from Opinion to Emotion Mining. ACM Computing Surveys (CSUR), 50(2):25:1-25:33, 2017.
5. Поздняков М.В., Осипов Н.А., Зудилова Т.В., Ананченко И.В., Иванов С.Е. Оценивание эмоциональной окраски текста при помощи нечёткой логики // Современные наукоёмкие технологии -2022. - № 10. - С. 32-36